Metabarcoding and Metagenomics 1: e14625
doi: 10.3897/mbmg.1.14625

OPEN ACCESS

Primer Validation

Short COI markers for freshwater macroinvertebrate metabarcoding

Ecaterina Edith Vamos‡, Vasco Elbrecht‡, Florian Leese‡

‡ University of Duisburg-Essen, Essen, Germany

Corresponding author: Ecaterina Edith Vamos (edith.vamos@uni-due.de), Vasco Elbrecht (vasco.elbrecht@uni-due.de)

Academic editor: Owen S. Wangensteen

Received: 21 Jun 2017 | Accepted: 10 Aug 2017 | Published: 20 Sep 2017

Citation: Vamos E, Elbrecht V, Leese F (2017) Short COI markers for freshwater macroinvertebrate metabarcoding. Metabarcoding and Metagenomics 1: e14625. https://doi.org/10.3897/mbmg.1.14625

Abstract

Species diversity of metazoan bulk samples can be rapidly assessed using cytochrome c oxidase I (COI) metabarcoding. However, in some applications often only degraded DNA is available, e.g. from poorly conserved museum specimens, environmental DNA (eDNA) filtered from water or gut content analyses. Here universal primer sets targeting only a short COI fragment are advantageous, as they often can still amplify short DNA fragments. Using PrimerMiner, we optimised two universal primer sets targeting freshwater macroinvertebrates based on NCBI and BOLD reference sequences. The fwh1 and fwh2 primer sets targeting a 178 and 205 bp region were tested *in vitro* by sequencing previously used freshwater macroinvertebrate mock communities as well as three monitoring samples from Romanian streams of unknown composition. They were further evaluated *in silico* for their suitability to amplify other insect groups. The fwh1 primer sets showed the most consistent amplification *in silico* and *in vitro*, detecting 92% of the taxa present in the mock communities, and allowing clear differentiation between the three macroinvertebrate communities from the Romanian streams. *In silico* analysis indicates that the short primers are likely to perform well even for non-freshwater insects. Comparing the performance of the new fwh1 primer sets to a highly degenerate primer set targeting a longer fragment (BF2+BR2) revealed that detection efficiency is slightly lower for the new primer set. Nevertheless, the shorter new primer pairs might be useful for studies that have to rely on degraded or poorly conserved DNA and thus be of importance for biomonitoring,

© Vamos E et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

conservation biological or molecular ecological studies. Furthermore, our study highlights the need for *in silico* evaluation of primer sets in order to detect design errors in primers (fwhR2) and find optimal universal primer sets for the target taxa of interest.

Keywords

metabarcoding, COI primers, degraded DNA, biodiversity assessment, freshwater macroinvertebrates, *in silico*

Introduction

Understanding ecosystem diversity and associated processes is essential for the management and protection of the biosphere. However, it is often challenging and time-consuming to reliably detect and identify organisms present in environmental samples (Haase et al. 2004). In freshwater ecosystems, for example, macroinvertebrates sampled for quality assessment often contain small organisms in immature life stages that can lack diagnostic morphological characters thus impeding species identification or even leading to misidentification (Sweeney et al. 2011). Here, DNA based specimen identification is a promising alternative to morphology-based identification methods. One of such technique is DNA metabarcoding where DNA is extracted from bulk samples (collected specimens) or environmental samples ("eDNA", e.g. filtered from water or sediment). Then PCR is used to amplify a barcoding gene, for animals usually the cytochrome c oxidase I (COI) region, followed by high-throughput sequencing (HTS) to generate a taxa inventory (Taberlet et al. 2012). This technique has already been applied to identify benthic biodiversity of freshwaters from bulk samples (e.g. Carew et al. 2013, Elbrecht et al. 2017b, Gibson et al. 2015, Hajibabaei et al. 2011) and eDNA (e.g. Deiner et al. 2016, Mächler et al. 2014, Bista et al. 2017), often in a water quality monitoring context. Nevertheless, metabarcoding is still a rather new approach and despite the significant progress made in recent years it still faces methodological as well as conceptual challenges (Elbrecht et al. 2017b, Leese et al. 2016). In particular, due to the high binding site variability in many metazoan groups, one issue is the design of appropriate universal primers (Sharma and Kobayashi 2014, Deagle et al. 2014, Elbrecht and Leese 2015). The proportion of taxa recovered with metabarcoding is dependent on the taxonomic resolution of the used gene marker (e.g. COI or ribosomal markers like 16S, Elbrecht et al. 2016), the length of the amplicon (Meusnier et al. 2008), universality of the primers and number of primer pairs used (Gibson et al. 2014) to amplify the taxonomic groups of interest (Elbrecht and Leese 2016, Elbrecht and Leese 2015, Deagle et al. 2014), as well as minor laboratory biases and stochastic effects (Leray and Knowlton 2017). For freshwater macrozoobenthos and most other metazoan species, usually primers targeting a short fragment of the standard COI barcoding region are used for metabarcoding, as this region shows a good taxonomic resolution (Hebert et al. 2003, Folmer et al. 1994). While the high variability of this region makes it possible to identify most taxa on species level, even when using a short ~150 bp fragment (Meusnier et al. 2008), it also makes it difficult to develop truly universal primer

sets (Sharma and Kobayashi 2014). Thus, the use of ribosomal markers that take advantage of the ribosomal stem regions has been suggested (Deagle et al. 2014), which are often well conserved across broad taxonomic groups. While ribosomal markers have been explored for freshwater taxa (Elbrecht et al. 2016) they likely offer no advantages in taxonomic resolution or taxa recovery compared to well-designed highly degenerated COI primer sets (Elbrecht and Leese 2017, Clarke et al. 2017). Additionally, barcoding gaps for the COI marker are well established for freshwater macroinvertebrates (Zhou et al. 2009, Zhou et al. 2010, Sweeney et al. 2011, Zhou et al. 2016) and available reference databases already cover most common freshwater taxa (Ratnasingham and Hebert 2007, Carew et al. 2017). Therefore, the good taxonomic resolution and already available reference data for the COI marker makes it an obvious choice for metabarcoding of freshwater macroinvertebrate communities. Recently, new universal primer sets specifically targeting freshwater macroinvertebrates were developed (BF+BR, Elbrecht and Leese 2017). In particular, the BF2+BR2 primer set that amplifies a 421 bp region of the COI Folmer fragment (Folmer et al. 1994) showed greatly reduced primer bias when tested with mock communities (Elbrecht and Leese 2017). Also on routine monitoring kick samples containing hundreds of morphologically identified freshwater specimens, this primer set recovered ~ 50 to 150% additional taxa while detecting a majority of the morphologically identified taxa (Elbrecht et al. 2017a, Elbrecht et al. 2017b). However, for amplification of degraded DNA e.g. from water samples (Barnes and Turner 2015), museum specimens (Shokralla et al. 2011) or for gut content analysis (Pompanon et al. 2011), targeting a shorter marker region of ~150 bp is assumed to increase amplification success (Herder et al. 2014, Thomsen and Willerslev 2015). The BF2+BR2 primer set is not expected to perform well on highly degraded DNA due to the long amplicon length. Further, while there are universal primers available that target only a short COI fragment, these often lack degeneracy and are developed for other taxonomic groups or ecosystems (Zeale et al. 2010, Meusnier et al. 2008).

In this study, we developed short metabarcoding primer pairs optimised to amplify degraded DNA from freshwater macroinvertebrates. We used COI reference sequences for 15 major freshwater groups important for bioassessment (see Elbrecht and Leese 2017 for details) to optimise base degeneracy for primers published by Folmer et al. (1994), Zeale et al. (2010), Leray et al. (2013) and Gibson et al. (2015). The short amplicons lead to fully overlapping paired end reads when sequenced on an Illumina MiSeq system, which is likely to increase the accuracy of the merged reads. The improved primer sets were tested using four macroinvertebrates mock communities each consisting of 52 freshwater taxa (Elbrecht and Leese 2015) as well as on three complete kick samples from Romanian streams. We also used the new primers to verify the correlation of biomass and sequence abundance within species as demonstrated in (Elbrecht and Leese 2015), in order to investigate if the same connection is found with highly degenerate primer sets. Additionally, we compared the novel primers *in silico* to a broader taxonomic range and alternative primers to explore their usefulness beyond the assessment of macroinvertebrate communities.

Material and Methods

Primer development

Two primer sets were developed using PrimerMiner (Elbrecht and Leese 2016) and a previously generated sequence alignment of 15 bioassessment relevant freshwater macroinvertebrate groups (Elbrecht and Leese 2017). The novel fwh1 and fwh2 primer sets amplify a short region of the cytochrome c oxidase I (COI) region of 178 and 205 bp in length respectively (Fig. 1, A). Both primer sets were based on primer sequences previously published in the literature (Table 1, Suppl. material 1), but primer degeneracy was increased to better match freshwater invertebrate taxa. For sequencing, the primers were ordered to include Illumina tails and individual inline barcodes for multiplex sequencing on the MiSeq system (Suppl. material 2, see Elbrecht and Leese 2015 for details on the “fusion primer” method). Using a 6 bp inline barcode for demultiplexing, the developed fusion primers can be used to individually tag up to 36 samples per primer set (Suppl. material 3).

Table 1.

COI primers developed in this study.

Primer name	Degenerated sequence (5'->3')	Direction	Based on
fwhF1	YTCHACWAAYCAYAARGAYATYGG	Forward	LCO1490 (Folmer et al. 1994)
fwhR1	ARTCARTTWCCRAAHCCHCC	Reverse	ZBJ-ArtR2c (Zeale et al. 2010)
fwhF2	GGDACWGGWTGAACWGTWTAYCCHCC	Forward	mlCOIintF (Leray et al. 2013)
fwhR2	GTRATWGCHCCDGCAARWACWGG	Reverse	ArR5 (Gibson et al. 2014)
fwhR2n	GTRATWGCHCCDGC**T**ARWACWGG	Reverse	ArR5 (Gibson et al. 2014)

In silico evaluation of primers

To explore the broader performance of the newly developed primers compared to the commonly used primers sets (Suppl. material 1), all primers were evaluated *in silico* for insect groups (following the taxonomy by Misof et al. 2014). Insect COI reference data was obtained in April 2016 and clustered into OTUs from NCBI and BOLD using PrimerMiner v0.3 as described previously (Elbrecht and Leese 2016, Elbrecht and Leese 2017). Sequence alignments were generated and used to evaluate the penalty scores given for primer mismatches using PrimerMiner v0.13 with the default settings (mm_position ="Position_v1", mm_type = "Type_v1"). Only orders with at least 100 OTUs were used to calculate the average penalty score for the respective primer.

Sample collection and processing

The performance of the fwh1 and fwh2 primer sets was evaluated using four previously used mock communities each containing 52 different freshwater taxa (sample A, B, C and

D) and one single species mock sample with 31 specimens with unique haplotypes and known biomass (Elbrecht and Leese 2015). Additionally, kick samples from three Romanian rivers (Almaşul, Călăţele and Valea Racilor, Suppl. material 4) were analyzed using both primer sets. The kick samples were collected in fall 2016, preserved in 95% ethanol and stored at -20°C for later molecular analysis. For the kick samples, no morphological identification of the macroinvertebrates was performed. Prior to DNA extraction, specimens were size sorted into small (S, body size < 2.5 x 5 mm), medium (M, up to 5 x 10 mm) and large (L, maximum size of 10 x 20 mm) specimens (Suppl. material 5, also see Elbrecht et al. 2017a).

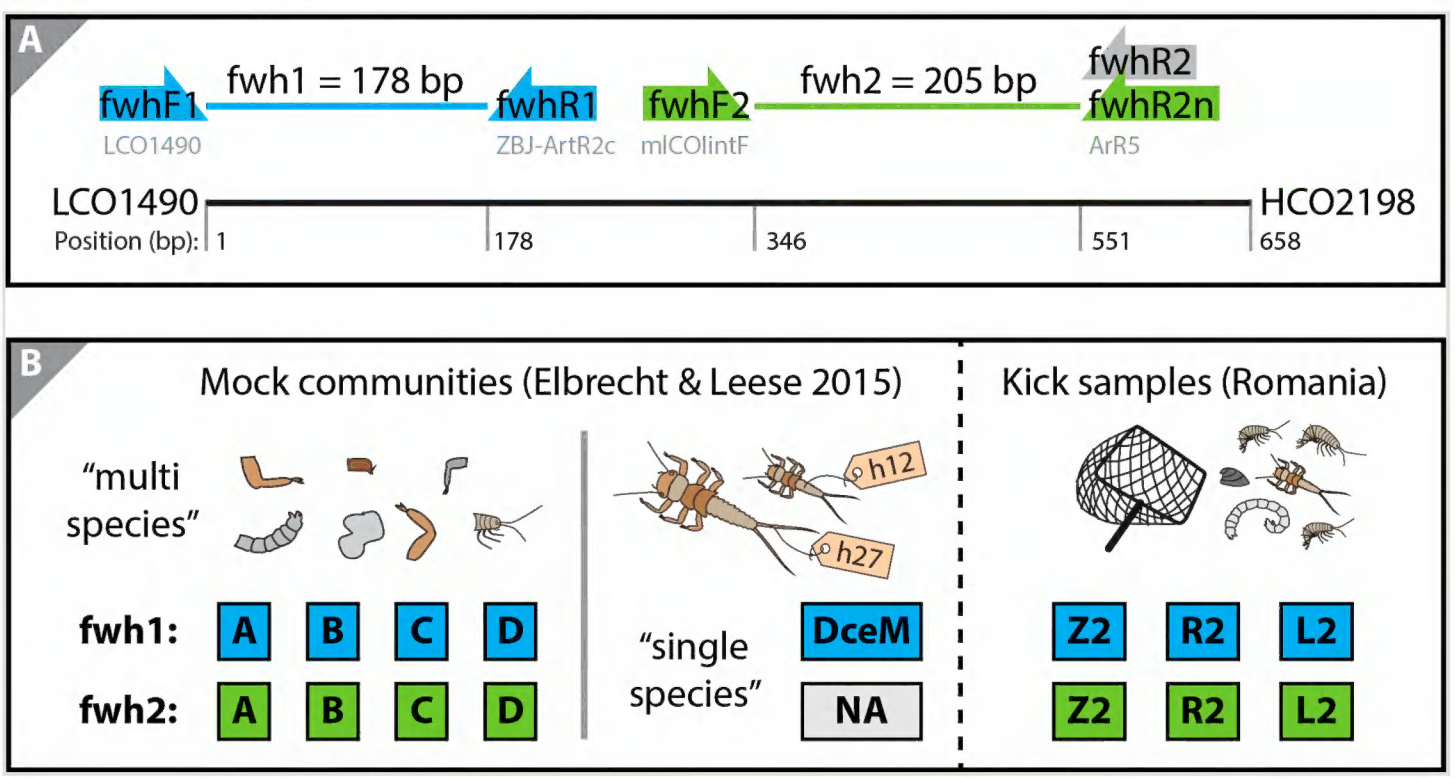

Figure 1.

Developed primer sets and samples sequenced for primer validation. Two independent PCR replicates were run and sequenced for each sample. **A**: Binding sites of the two primer sets (fwhF1+fwhR1 and fwhF2+fwhR2) targeting a 178 and 205 bp fragment internal to the COI Folmer barcoding region (Folmer et al. 1994). The fwhR2 primer was affected by a design error, thus an improved version (fwhR2n) was developed. **B**: Overview of the sequenced benthic communities and mock samples to test and validate the developed primer sets. Five mock communities (four multi and one single species) from Elbrecht and Leese (2015), as well as three kick samples collected from streams in Romania (Călățele River: L2, Almaşul River: R2, Valea Racilor River: Z2), were collected and tested using the fwh1 and fwh2 primer sets (except for sample DceM that could only be amplified using the fwh1 primer set).

DNA extraction and tissue pooling

Specimens of each size category (S, M & L) were dried overnight in sterile Petri dishes to remove the ethanol. Specimens from each category were homogenised using an IKA ULTRA-TURRAX Tube Drive control system (IKA, Staufen, Germany) with sterile 20 mL tubes and 10 steel beads (5 mm Ø) by grinding at 4000 rpm for 30 minutes. Approximately equal amounts of grinded tissue from each category were digested following a modified

salt DNA extraction protocol (on average 13.41 mg of tissue, SD = 12.34 mg, Sunnucks and Hales 1996, Elbrecht et al. 2017a). Next, the lysate was pooled proportionately to the abundance of individuals in each size category to reduce the overrepresentation of large specimens (see Elbrecht et al. 2017a for details). Further, 20 µl of the extracted DNA from each respective sample was digested with 1 µL RNase A (10 mg/mL, Life Technologies, Darmstadt, Germany) and cleaned up using a MinElute Reaction Cleanup Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. DNA concentrations were quantified fluorometrically using Qubit fluorometer (HS Kit, ThermoFisher Scientific, MA, USA) and concentrations for all samples were adjusted to 25 ng/µL for PCR.

DNA metabarcoding and bioinformatics

The five mock communities and three kick samples were amplified in duplicates in a one-step PCR using the developed freshwater primers (Table 1). Suppl. material 6 gives an overview of fusion primer combinations used to uniquely tag each sample. Each PCR reaction was composed of 1× Multiplex PCR Master Mix (Qiagen Multiplex PCR Plus Kit, Qiagen, Germany) 0.5 µM of each primer, 25 ng DNA, filled up with HPLC H_2O (Carl Roth GmbH, Karlsruhe, Germany) to a total volume of 50 µL. PCR reactions were run in a Biometra TAdvanced Thermocycler (Biometra, Göttingen, Germany) using the following program 95°C for 5 min, 34 cycles of 95°C for 30 s, 52°C (for the fwhF1+fwhR1 primer pair) or 58°C (for the fwhF2+fwhR2 primer pair) for 30 s, 72°C for 2 min, and 72°C for 10 min. The annealing temperatures for both primer sets were established by first running a gradient PCR on DNA from the multi species mock communities (gradient temperature 43.7 - 70.3°C, Suppl. material 7). The annealing temperatures for the respective primer pair were chosen a few degrees below the temperature of the last visible band, to ensure efficient and consistent amplification. PCR products from the one-step PCR were purified and left size selected using SPRIselect (Beckman Coulter, CA, USA) with a ratio of 0.76x and the DNA concentration was quantified with a Qubit fluorometer, High Sensitivity Kit (Thermo Fisher Scientific, MA, USA) and Fragment Analyzer Automated CE System using NGS Standard Sensitivity kit (Advanced Analytical, Heidelberg, Germany). The mean DNA concentration from both measurements were used to pool PCR products by equal molarity. This final library was additionally purified with the MinElute Reaction Cleanup Kit (Qiagen, Hilden, Germany), as a precaution due to BSA interfering with the PCR clean-up using SPRIselect (Elbrecht et al. 2017a). Sequencing was done on two runs of an Illumina MiSeq system using a 250 bp paired end read kit (v2) and 5% PhiX spike-in. Sequencing was carried out by GATC Biotech GmbH (Konstanz, Germany). Raw sequence data were processed using a modified version of the UPARSE pipeline (Edgar 2013, v9.0.2132), which is available at GitHub (JAMP version 0.17 - https://github.com/VascoElbrecht/JAMP/). The exact commands run to reproduce the analysis are available in Suppl. material 8. In short, reads were demultiplexed, paired-end merged using usearch, reverse complement sequences generated where necessary, quality filtered (maxee = 0.05) and pre-processed (primer removal, Cutadapt v1.9 (Martin 2011), discarding of reads +/- 10 bp of the expected length, dereplication with removing singletons; minsize = 2). Before applying clustering (97% similarity) all retained sequences for the A, B, C and D samples were pooled and the three Romanian samples were also pooled. Reads including

singletons were remapped against the OTUs and clusters with at least 0.003% abundance in one sample retained (both replicates). OTUs were identified using sequences from previous studies as references and comparison against BOLD and NCBI reference databases with JAMP. For the single species mock samples (DceM) filtered dereplicated reads of exact 178 bp length were directly mapped against the expected haplotypes (Suppl. material 8) and all matching hits with at least 0.003% abundance retained.

Results

Primer design and *in silico* evaluation

Two primer sets were developed targeting short COI fragment lengths of 178 bp and 205 bp respectively (Fig. 1). *In silico* evaluation of the developed primer sets on insect orders was only carried out when preparing this manuscript, and it became evident that the fwhR2 primer had a design flaw (Fig. 2). At position 9 from the 3' end (Table 1), an Adenine (A) was used instead of a Thymine (T), leading to a poor estimated primer performance (mean penalty score of 144.7). This mistake was corrected afterwards in the fwh2n version of this primer (Table 1), which shows a decreased average penalty score (53.4). However, the new improved version of the primer was only tested *in silico* and all laboratory tests were carried out using the flawed fwhR2 version of this primer.

The other evaluated primer sets showed mixed performances depending on the degeneracy of the respective primers. A lack of degeneracy resulted in rather high penalty scores, as was the case for the LCO1490+HCO2198 and ZBJ-ArtF1c+ZBJ-ArtR2c primer sets (scores above 100, Zeale et al. 2010, Folmer et al. 1994). Primers incorporating an Inosine, e.g. ArF5+ArR5 (Gibson et al. 2014), or a high degeneracy, e.g. the BF+BR primer sets (Elbrecht and Leese 2017), showed low average penalty scores (below 40). The universal BF+BR and mlCOIintF primers showed increased penalty scores for a few groups that have more variable primer binding regions in the template DNA (Thysanoptera, Phasmatodea or Raphidioptera). Some of the primers binding at the LCO1490 binding site showed high scores due to misaligned sequences or low number of OTUs. The newly designed primers fwhF1, fwhF2 and fwhR1 had lower penalty scores than the primer sets they are based on, while the fwhR2n primer set showed a higher penalty score (53.4) than the ArR5 primer set with a score of 6.9.

Metabarcoding and Illumina sequencing

Both fwh primer sets successfully amplified the four multispecies mock communities (A, B, C and D) as well as the three Romanian stream kick samples. The fwh2 primers only produced a weak amplicon band on the agarose gel for the DceM sample, which was therefore only sequenced using the fwh1 primer set. Both Illumina MiSeq runs were successful for all sequenced samples with an average number of 1.40 (fwh1) and 0.74 (fwh2) million sequences obtained for each replicate (SD = 0.26 and 0.13, Suppl. material 9). Raw sequencing data are available on NCBI SRA (SRR5295658 and SRR5295659).

OTU tables including assigned taxonomy and OTU sequences are available as supporting information (Suppl. material 10).

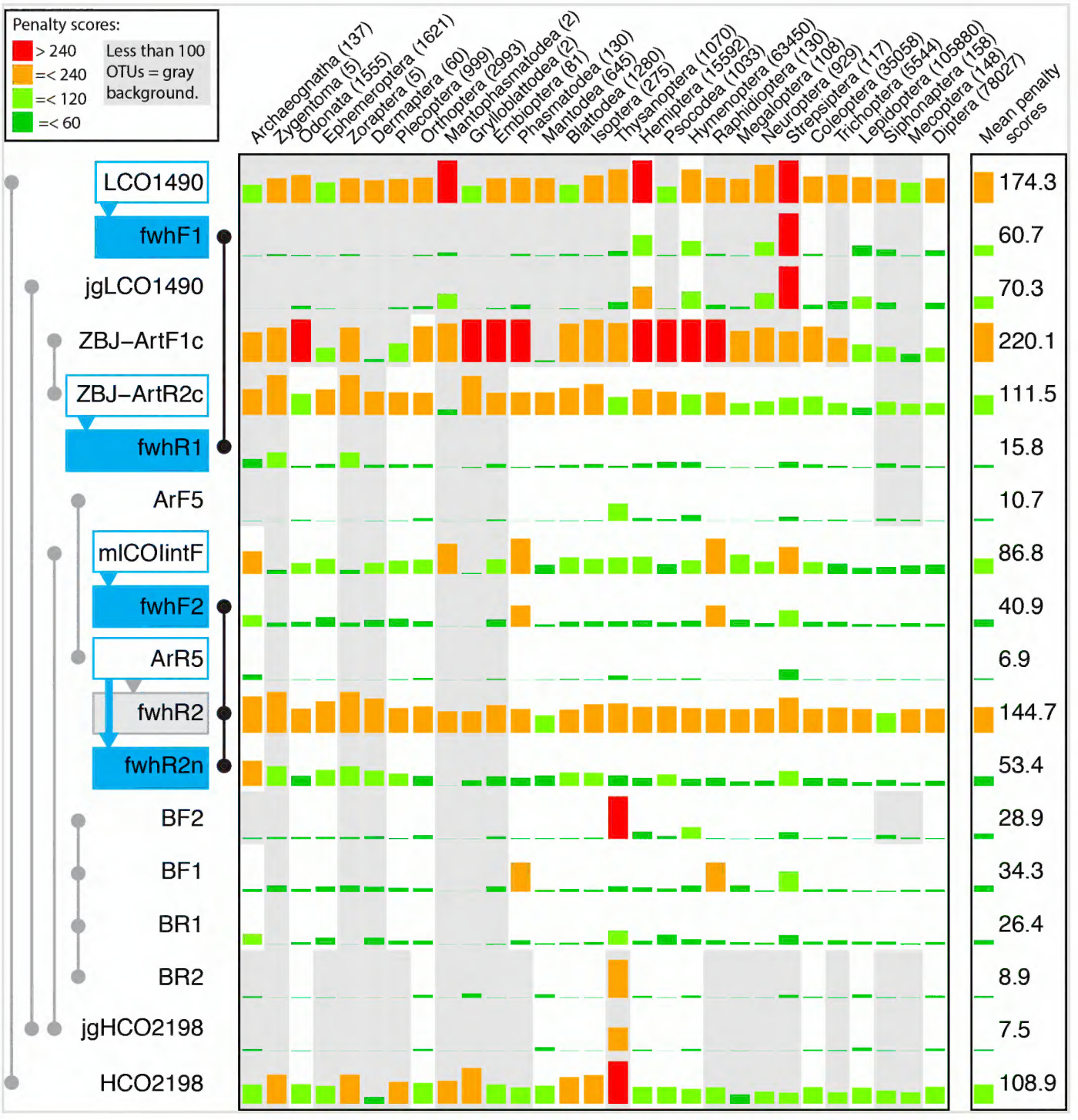

Figure 2.

In silico evaluation of insect groups (after Misof et al. 2014) for selected metabarcoding primer pairs. COI reference sequences for primer evaluation were obtained from BOLD and NCBI using PrimerMiner and processed into OTUs (3% similarity). For primer-template mismatches, penalty scores were calculated using PrimerMiner (lower penalty score = better expected primer performance). The individual mean penalty scores are given in bar plots for each primer and insect order. The average penalty score was calculated for each primer for orders with at least 100 OTUs for the respective primer pair. The typically used primer combinations are indicated by connected grey lines on the left, as well as black lines for the newly developed primer pairs. Templates for primer development are indicated with blue boxes, while the newly developed primers are highlighted with blue backgrounds. The fwhR2 primer had a design error and is highlighted in grey.

Taxa recovery in mock and bulk samples

For the four mock communities, most of the taxa were recovered by both primer sets. While fwh1 primers detected 48 taxa out of 52, the fwh2 performed poorer, recovering 46 taxa (Fig. 3). Variation in logarithmic taxa read abundance was much lower for the fwh1 amplicons (SD = 0.62) than for the fwh2 primer set (SD = 0.97) across the mock community samples (Table 2). The fwh1 primer set also showed the highest precision (deviation from expected read abundance). For DceM mock community, only the fwh1 primer set produced an amplicon, as the fwh2 primer set did not amplify Perlidae efficiently (see also Fig. 3). Because the fwh1 fragment is shorter than the previously sequenced Folmer COI fragment (Elbrecht and Leese 2015), only 15 of the original 31 haplotypes could possibly be distinguished (Fig. 4). All 15 expected haplotypes in the DceM community were recovered with the fwh1 primer set. Both PCR replicates showed the same trend in the relative sequence abundance with an expected ratio of relative haplotype abundance approximately equal to 1 (Fig. 4, B). Also, both replicates had very similar read composition, with only rare reads being unique to specific samples (Suppl. material 11).

Table 2.

Number of morphotaxa recovered with the fwh and Folmer primers from previous tests (Elbrecht and Leese 2015).

Taxonomic group	No. of specimens	No. of specimens recovered with specific primer combination		
		LCO1490+HCO2198	fwhF1+fwhR1	fwhF2+fwhR2
Ephemeroptera	8	7 (88%)	8 (100%)	7 (88%)
Plecoptera	4	4 (100%)	4 (100%)	4 (100%)
Trichoptera	15	13 (86%)	14 (93%)	15 (100%)
Diptera	8	7 (88%)	8 (100%)	8 (100%)
Other insects	7	7 (100%)	7 (100%)	7 (100%)
Other metazoan	10	5 (50%)	7 (70%)	5 (50%)
Σ All insects	42	38 (91%)	41 (98%)	41 (98%)
SD *		1.01	0.62	0.97
Precision **		0.72	0.43	0.68
Σ All taxa	52	43 (83%)	48 (92%)	46 (88%)

* Mean standard deviation (SD) of $\log_{10}$ read abundance from each insect taxon that was detected (specimens with < 0.003% read abundance discarded). **Precision defined as the SD of the mean $\log_{10}$ distance to the expected read abundance, calculated for each morphotaxon (all taxa).

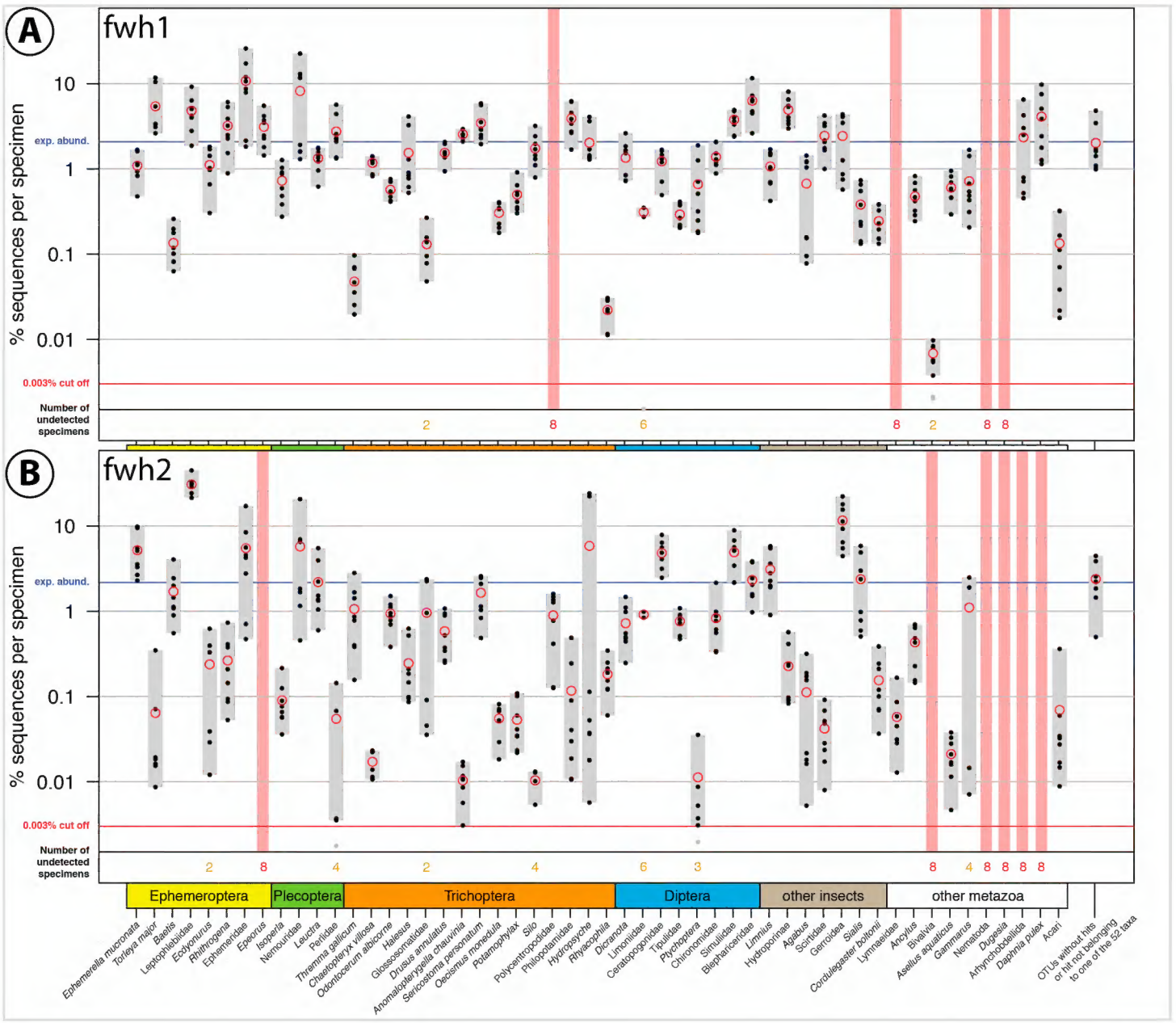

Figure 3.

Comparison of fwh1 (A) and fwh2 (B) primer performance, both tested with the same four bulk samples with two independent PCR replicates for each sample. Each respective sample contained 52 morphologically distinct macroinvertebrate taxa ("TierMix": A, B, C & D). The 52 taxa are shown on the x-axis with the number of reads obtained for each morphotaxon indicated by black dots on the logarithmic y-axis (mean relative abundance of detected morphotaxa is indicated by red circles, replicates are plotted). Sequence abundance was normalised across the samples and the amount of tissue used in each DNA extraction. Only OTUs which had a minimum abundance of 0.003% in at least one of the four samples were included in the analysis. Number of samples for which a morphotaxon was not detected is indicated by orange and red numbers in each plot. A thick vertical line in light red indicates if a morphotaxon was not detected.

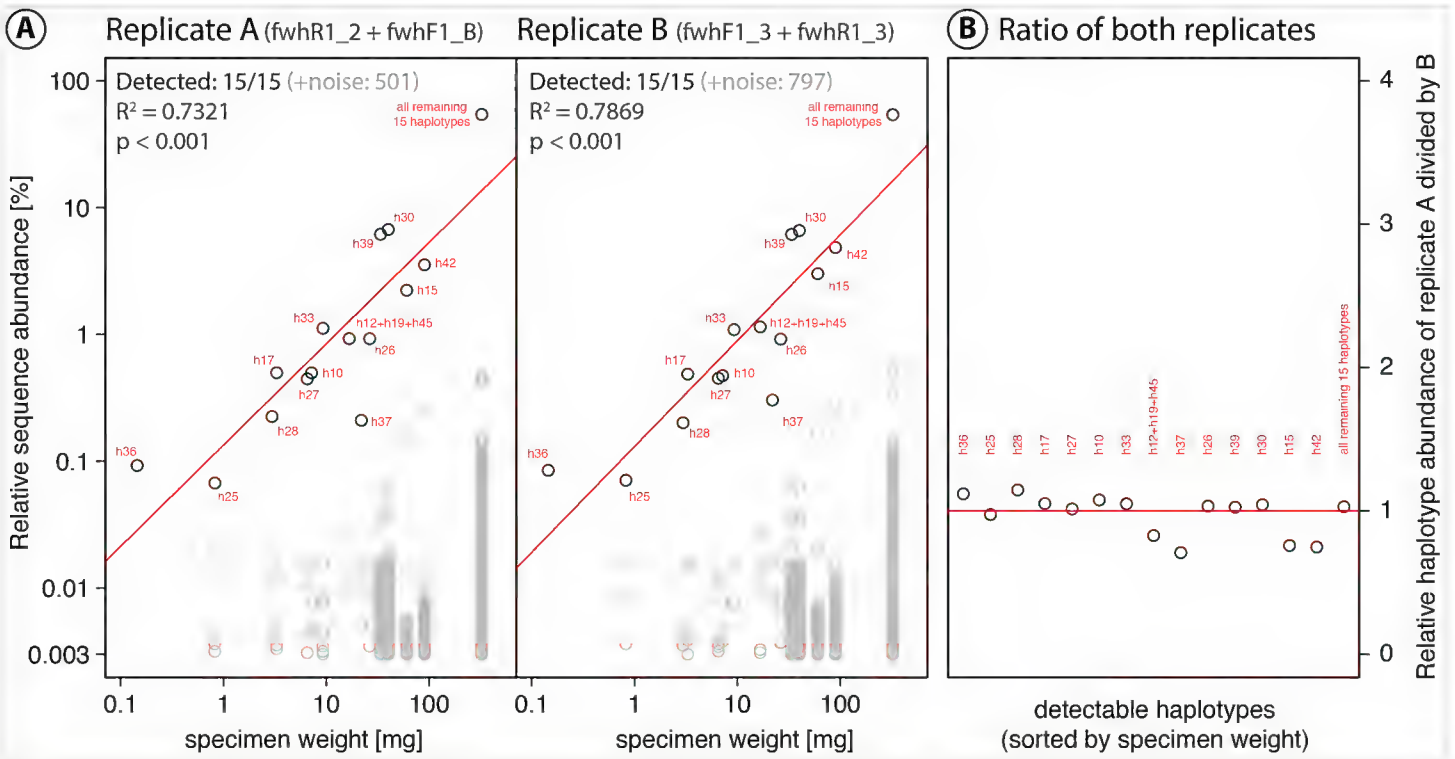
A
Replicate A (fwhR1_2 + fwhF1_B)
Replicate B (fwhF1_3 + fwhR1_3)
B Ratio of both replicates
Detected: 15/15 (+noise: 501)
R² = 0.7321
p < 0.001
Detected: 15/15 (+noise: 797)
R² = 0.7869
p < 0.001
all remaining 15 haplotypes
Relative sequence abundance [%]
specimen weight [mg]
detectable haplotypes (sorted by specimen weight)
Relative haplotype abundance of replicate A divided by B

Figure 4.

Detection of haplotypes in the tested single species mock community (DceM) using the fwh1 primer set. Sequences below 0.003% relative read abundance were discarded. **A**: Relative abundance of detected haplotypes in both PCR replicates plotted against cumulative specimen weight (red line indicates linear regression). Because the fwh1 fragment is shorter than the previously sequenced Folmer COI fragment (Elbrecht and Leese 2015), only a maximum of 15 haplotypes can be detected with the short COI fragment. **B**: Ratio of relative haplotype abundance when dividing replicate A by replicate B with a red line indicating the expected value of 1.

For the three Romanian samples the ecological quality state of the rivers was assessed only on the expert judgment (visual assessment, Suppl. material 4) and not based on a standardised assessment using morphologically identified macroinvertebrate taxa from kick samples (see Suppl. material 5 for pictures of samples composition). However, by analyzing the diversity of EPT taxa (Ephemeroptera, Plecoptera, Trichoptera, typically highly pollution-sensitive taxa), it is possible to get a proxy for the ecological condition of the streams. For the study sites L2 and Z2 (good to mediocre ecological status according to expert judgment) 42.66% and 46.44% of the OTUs were identified as EPT, while at the R2 site (poor ecological state) only 7.82% EPT taxa were detected (Suppl. material 12). For fwh2 primer set we obtained very similar results; for Z2 and L2 sites the EPT is represented by 18.53% and 29.23%, while for R2 site 8.47% of the OTUs were assigned to EPT taxa. Taxonomic richness of the streams communities is in good agreement with their ecological state. Our primer pairs also amplified non-target species, with high identity matches (>= 97%) to the reference databases, such as hop aphids, moths and few freshwater fish species (e.g. gudgeon, minnows and stone loaches). The principal component analysis of the macroinvertebrate OTUs obtained from the fwh1 and fwh2 primer sets showed clear differentiation between the three Romanian samples, while consistently grouping PCR replicates of the same sites together (Fig. 5).

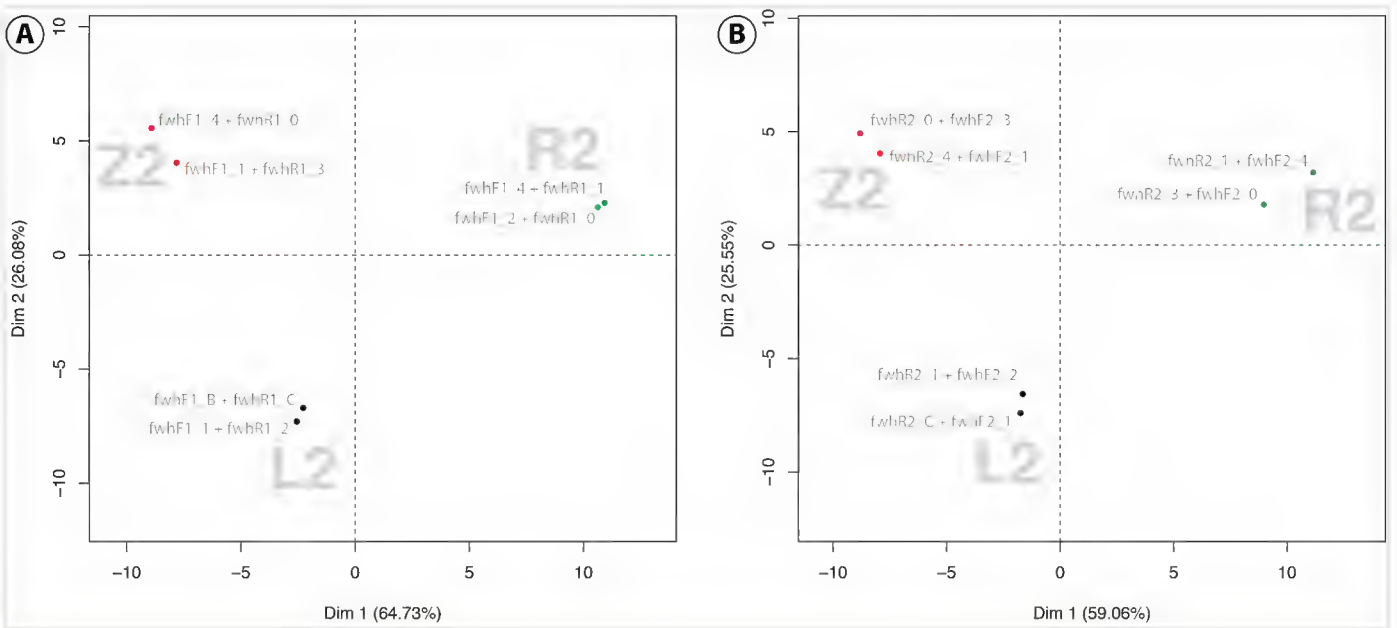

Figure 5.

Principal component analysis (PCA) of freshwater macroinvertebrate OTUs detected with the fwh1 (**A**) and fwh2 (**B**) primer set in the three Romanian river samples. PCR replicates of the identical samples are shown with the same colour.

Discussion

Primer development and performance

Using PrimerMiner we have developed two short universal metabarcoding primer sets targeting freshwater macroinvertebrates. As previously reported, a short 150 bp barcode marker is sufficient to identify most insect taxa on species level (Meusnier et al. 2008). Also, PCR with short amplicons is expected to work better when dealing with highly degraded DNA (Dalvin et al. 2010, Mitchell 2015, Schäffer et al. 2017). Additionally, in contrast to previously developed longer universal markers like the BF2+BR2 primer set (Elbrecht and Leese 2017), fragments of ~200 bp length can be paired-end sequenced on the Illumina NextSeq system increasing throughput ten-fold compared to the MiSeq/HiSeq system, which is commonly used for amplicon sequencing (e.g. Schöfl et al. 2017). The MiniSeq system can also be used if only a few samples have to be sequenced. Additionally, completely overlapping amplicons can reduce sequencing errors when paired-end merged (Kozich et al. 2013, Eren et al. 2013).

In silico evaluation of the newly developed primer sets revealed that all of them showed low penalty scores, except for the fwh2R primer, where a design error was introduced causing mismatches across most taxa. Unfortunately, *in silico* evaluation was only carried out when preparing this manuscript, therefore a corrected version of the primer (fwhR2n) could not be tested *in vitro* with mock and kick samples. Compared to the fwh1 primer set, the fwhR2n primer still shows a high average penalty score of above 50, thus even the improved reverse primer version is not likely to perform particularly well for amplification of insects. The *in silico* evaluation also showed that the fwh1, BF+BR as well as the ArF5

+ArR5 primer sets (Gibson et al. 2014, Elbrecht and Leese 2017) are likely to work well across most insect orders.

Both the fwh1 and the fwh2 primer sets (using the flawed fwhR2 primer) were additionally tested on several mock communities as well as complete kick samples. While both primer sets could clearly differentiate the three stream sites in a principal component analysis, the fwh2 primer set showed a higher primer bias when amplifying mock communities, each containing 52 freshwater macroinvertebrates. The primer set also failed to amplify the DceM mock community containing specimens of a single stonefly and was thus not included in sequencing. Perlidae specimens were also underrepresented in the multi species mock community with the fwh2 primer set, indicating strong primer bias for this group. Both the fwh1 and fwh2 primer sets detected 98% of the freshwater insects present in the multi species mock communities. While both universal primer sets show higher detection rates and reduced primer bias compared to the standard COI Folmer primer sets (Folmer et al. 1994, Elbrecht and Leese 2015), the recently developed BF2+BR2 primer set for a longer fragment shows less primer bias than the here proposed novel primer sets (Elbrecht and Leese 2017). Therefore, as long as DNA degradation is not a concern we recommend the use of the BF2+BR2 primer sets for DNA metabarcoding of freshwater macroinvertebrate and insect samples (except for the derived Thysanoptera group). The ArF5+ArR5 primer set also performed well in the *in silico* evaluation but should be further validated *in vitro*. Additionally, when dealing with highly degraded DNA, especially the fwh1 primer set might prove to be useful.

Validation of sequence / biomass relationships within species

We also used the new fwh1 primer set to test for a linear relationship of sequence abundance to specimen biomass within species, which we previously explored for the Folmer primer sets (Folmer et al. 1994, Elbrecht and Leese 2015). In the previous study, a single mock sample containing 31 unique haplotypes of a single stonefly species was amplified and a significant linear relationship between numbers of sequences and specimen biomass was detected. While this sounds promising for estimating taxa abundance or biomass from metabarcoding data, reliable estimates are difficult to obtain due to often severe primer bias between different species (Piñol et al. 2014, Elbrecht and Leese 2015). However, the sequence / biomass relationship within species was tested here again using the shorter fwh1 primer set that has higher degeneracy and can be paired end merged with a complete overlap of forward and reverse sequencing reads, potentially reducing sequencing errors compared to the previous study where the standard Folmer primers were used (Folmer et al. 1994). As expected, with the fwh1 primer set also a significant correlation between read abundance and specimen biomass was detected. In addition, by comparing the two sequenced PCR replicates, we could analyse the false positive haplotypes generated by sequencing errors and chimeras. In particular abundant haplotypes showed hundreds of artificial haplotypes, likely derived from sequencing errors and chimera formation. These errors were consistent between PCR replicates, with only low abundant sequences being unique to the respective sample (Lange et al. 2015). Both PCR replicates showed differences in low abundant reads, which were likely to be

generated by sequencing errors or chimeric sequences. However, the majority of the false positive haplotypes were shared between both replicates indicating systematic origins, for example, chimera formation or sequencing errors on high abundant haplotypes. Amplification of abundant taxa is typically very consistent, with stochastic effects mostly affecting low abundant taxa / sequences as also demonstrated in other studies (Leray and Knowlton 2017), highlighting that the use of PCR replicates might not substantially increase the reliability of DNA metabarcoding results (Smith and Peay 2014). Some of this false positives, however, might also be present due to mitochondrial heteroplasmy or the presence of 'numts', i.e. nuclear sequences of mitochondrial origin (Bensasson 2001).

Conclusions

DNA metabarcoding is a powerful tool for understanding and assessing aquatic biodiversity. While there are well-designed and evaluated primer sets available to generate comparatively long amplicons (BF2+BR2), these might fail when targeting samples of highly degraded DNA. The primer sets developed here are suggested as a valuable alternative for such special cases where longer fragments are difficult to obtain. Our primer evaluation, especially of the fwh1 primer set, demonstrates the excellent performance with mock samples and the ability to clearly differentiate between the complete freshwater invertebrate communities from three Romanian streams. We therefore encourage the application of the fwh1 primer set for gut content analysis, poorly conserved museum specimens and when targeting highly degraded environmental DNA from e.g. water or soil samples.

Acknowledgements

We would like to thank Karina Battes from Babes-Bolyai University, Cluj-Napoca for kindly providing the kick samples. EEV was supported by a fellowship of the Dr. Musat V. Bodnarescu foundation. We would further like to thank the leeselab journal club proofreading this manuscript. This publication is part of the EU COST Action *DNAqua-Net* (CA15219).

Author contributions

VE, EEV and FL conceived the ideas and designed methodology; EEV carried out the laboratory work; VE performed bioinformatic analyses; VE and EEV led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Conflicts of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

- Barnes M, Turner C (2015) The ecology of environmental DNA and implications for conservation genetics. Conservation Genetics 17 (1): 1-17. https://doi.org/10.1007/s10592-015-0775-4
- Bensasson D (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends in Ecology & Evolution 16 (6): 314-321. https://doi.org/10.1016/s0169-5347(01)02151-6
- Bista I, Carvalho G, Walsh K, Seymour M, Hajibabaei M, Lallias D, Christmas M, Creer S (2017) Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. Nature Communications 8: 14087. https://doi.org/10.1038/ncomms14087
- Carew ME, Pettigrove VJ, Metzeling L, Hoffmann AA (2013) Environmental monitoring using next generation sequencing: rapid identification of macroinvertebrate bioindicator species. Frontiers in Zoology 10 (1): 45. https://doi.org/10.1186/1742-9994-10-45
- Carew ME, Nichols SJ, Batovska J, Clair RS, Murphy NP, Blacket MJ, Shackleton ME (2017) A DNA barcode database of Australia's freshwater macroinvertebrate fauna. Marine and Freshwater Research https://doi.org/10.1071/mf16304
- Clarke L, Beard J, Swadling K, Deagle B (2017) Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. Ecology and Evolution 7 (3): 873-883. https://doi.org/10.1002/ece3.2667
- Dalvin S, Glover KA, Sørvik AG, Seliussen BB, Taggart JB (2010) Forensic identification of severely degraded Atlantic salmon (Salmo salar) and rainbow trout (Oncorhynchus mykiss) tissues. Investigative Genetics 1 (1): 12. https://doi.org/10.1186/2041-2223-1-12
- Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P (2014) DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. Biology Letters 10 (9): 20140562-20140562. https://doi.org/10.1098/rsbl.2014.0562
- Deiner K, Fronhofer EA, Mächler E, Walser J, Altermatt F (2016) Environmental DNA reveals that rivers are conveyer belts of biodiversity information. Nature communications 7: 12544. https://doi.org/10.1038/ncomms12544
- Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nature Methods 10 (10): 996-998. https://doi.org/10.1038/nmeth.2604
- Elbrecht V, Leese F (2015) Can DNA-Based Ecosystem Assessments Quantify Species Abundance? Testing Primer Bias and Biomass—Sequence Relationships with an Innovative Metabarcoding Protocol. PLOS ONE 10 (7): e0130324. https://doi.org/10.1371/journal.pone.0130324
- Elbrecht V, Leese F (2016) PrimerMiner: an rpackage for development and in silicovalidation of DNA metabarcoding primers. Methods in Ecology and Evolution 8 (5): 622-626. https://doi.org/10.1111/2041-210x.12687

- Elbrecht V, Leese F (2017) Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. Frontiers in Environmental Science 5 https://doi.org/10.3389/fenvs.2017.00011
- Elbrecht V, Peinert B, Leese F (2017a) Sorting things out - assessing effects of unequal specimen biomass on DNA metabarcoding. Ecology and Evolution https://doi.org/10.1002/ece3.3192
- Elbrecht V, Vamos EE, Meissner K, Aroviita J, Leese F (2017b) Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. Methods in Ecology and Evolution https://doi.org/10.1111/2041-210x.12789
- Elbrecht V, Taberlet P, Dejean T, Valentini A, Usseglio-Polatera P, Beisel J, Coissac E, Boyer F, Leese F (2016) Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. PeerJ 4: e1966. https://doi.org/10.7717/peerj.1966
- Eren AM, Vineis J, Morrison H, Sogin M (2013) A Filtering Method to Generate High Quality Short Reads Using Illumina Paired-End Technology. PLoS ONE 8 (6): e66643. https://doi.org/10.1371/journal.pone.0066643
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular marine biology and biotechnology 3 (5): 294-9.
- Gibson J, Shokralla S, Curry C, Baird D, Monk W, King I, Hajibabaei M (2015) Large-Scale Biomonitoring of Remote and Threatened Ecosystems via High-Throughput Sequencing. PLOS ONE 10 (10): e0138432. https://doi.org/10.1371/journal.pone.0138432
- Gibson J, Shokralla S, Porter TM, King I, Konynenburg Sv, Janzen DH, Hallwachs W, Hajibabaei M (2014) Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. Proceedings of the National Academy of Sciences 111 (22): 8007-8012. https://doi.org/10.1073/pnas.1406468111
- Haase P, Lohse S, Pauls S, Schindehütte K, Sundermann A, Rolauffs P, Hering D (2004) Assessing streams in Germany with benthic invertebrates: development of a practical standardised protocol for macroinvertebrate sampling and sorting. Limnologica - Ecology and Management of Inland Waters 34 (4): 349-365. https://doi.org/10.1016/s0075-9511(04)80005-7
- Hajibabaei M, Shokralla S, Zhou X, Singer GC, Baird D (2011) Environmental Barcoding: A Next-Generation Sequencing Approach for Biomonitoring Applications Using River Benthos. PLoS ONE 6 (4): e17497. https://doi.org/10.1371/journal.pone.0017497
- Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences 270 (1512): 313-321. https://doi.org/10.1098/rspb.2002.2218
- Herder J, Valentini A, Bellemain E, Dejean T, Delft Jv, Thomsen PF, Taberlet P (2014) Environmental DNA - a review of the possible applications for the detection of (invasive) species. Stichting RAVON
- Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. Applied and Environmental Microbiology 79 (17): 5112-5120. https://doi.org/10.1128/aem.01043-13

- Lange A, Jost S, Heider D, Bock C, Budeus B, Schilling E, Strittmatter A, Boenigk J, Hoffmann D (2015) AmpliconDuo: A Split-Sample Filtering Protocol for High-Throughput Amplicon Sequencing of Microbial Communities. PloS one 10 (11): e0141590. https://doi.org/10.1371/journal.pone.0141590
- Leese F, Altermatt F, Mächler E, Ekrem T, Bouchez A, et al. (2016) DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. Pensoft Publishers https://doi.org/10.5167/UZH-132075
- Leray M, Knowlton N (2017) Random sampling causes the low reproducibility of rare eukaryotic OTUs in Illumina COI metabarcoding. PeerJ 5: e3006. https://doi.org/10.7717/peerj.3006
- Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Frontiers in Zoology 10: 43-43.
- Mächler E, Deiner K, Steinmann P, Altermatt F (2014) Utility of environmental DNA for monitoring rare and indicator macroinvertebrate species. Freshwater Science 33 (4): 1174-1183. https://doi.org/10.1086/678128
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17 (1): 10. https://doi.org/10.14806/ej.17.1.200
- Meusnier I, Singer GAC, Landry J, Hickey DA, Hebert PDN, Hajibabaei M (2008) A universal DNA mini-barcode for biodiversity analysis. BMC genomics 9: 214. https://doi.org/10.1186/1471-2164-9-214
- Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, Niehuis O, Petersen M, Izquierdo-Carrasco F, Wappler T, Rust J, Aberer AJ, Aspöck U, Aspöck H, Bartel D, Blanke A, Berger S, Böhm A, Buckley TR, Calcott B, Chen J, Friedrich F, Fukui M, Fujita M, Greve C, Grobe P, Gu S, Huang Y, Jermiin LS, Kawahara AY, Krogmann L, Kubiak M, Lanfear R, Letsch H, Li Y, Li Z, Li J, Lu H, Machida R, Mashimo Y, Kapli P, McKenna DD, Meng G, Nakagaki Y, Navarrete-Heredia JL, Ott M, Ou Y, Pass G, Podsiadlowski L, Pohl H, von Reumont BM, Schütte K, Sekiya K, Shimizu S, Slipinski A, Stamatakis A, Song W, Su X, Szucsich NU, Tan M, Tan X, Tang M, Tang J, Timelthaler G, Tomizuka S, Trautwein M, Tong X, Uchifune T, Walzl MG, Wiegmann BM, Wilbrandt J, Wipfler B, Wong TKF, Wu Q, Wu G, Xie Y, Yang S, Yang Q, Yeates DK, Yoshizawa K, Zhang Q, Zhang R, Zhang W, Zhang Y, Zhao J, Zhou C, Zhou L, Ziesmann T, Zou S, Li Y, Xu X, Zhang Y, Yang H, Wang J, Wang J, Kjer KM, Zhou X (2014) Phylogenomics resolves the timing and pattern of insect evolution. Science (New York, N.Y.) 346 (6210): 763-7. https://doi.org/10.1126/science.1257570
- Mitchell A (2015) Collecting in collections: a PCR strategy and primer set for DNA barcoding of decades-old dried museum specimens. Molecular Ecology Resources 15 (5): 1102-1111. https://doi.org/10.1111/1755-0998.12380
- Piñol J, Mir G, Gomez-Polo P, Agustí N (2014) Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. Molecular Ecology Resources 15 (4): 819-830. https://doi.org/10.1111/1755-0998.12355
- Pompanon F, Deagle BE, Symondson WOC, Brown DS, Jarman SN, Taberlet P (2011) Who is eating what: diet assessment using next generation sequencing. Molecular ecology 21 (8): 1931-50. https://doi.org/10.1111/j.1365-294X.2011.05403.x

- Ratnasingham S, Hebert PDN (2007) bold: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular ecology notes 7 (3): 355-364. https://doi.org/10.1111/j.1471-8286.2007.01678.x
- Schäffer S, Zachos F, Koblmüller S (2017) Opening the treasure chest: A DNA-barcoding primer set for most higher taxa of Central European birds and mammals from museum collections. PLOS ONE 12 (3): e0174449. https://doi.org/10.1371/journal.pone.0174449
- Schöfl G, Lang K, Quenzel P, Böhme I, Sauter J, Hofmann J, Pingel J, Schmidt A, Lange V (2017) 2.7 million samples genotyped for HLA by next generation sequencing: lessons learned. BMC Genomics 18 (1): . https://doi.org/10.1186/s12864-017-3575-z
- Sharma P, Kobayashi T (2014) Are “universal” DNA primers really universal? Journal of Applied Genetics 55 (4): 485-496. https://doi.org/10.1007/s13353-014-0218-9
- Shokralla S, Zhou X, Janzen D, Hallwachs W, Landry J, Jacobus L, Hajibabaei M (2011) Pyrosequencing for Mini-Barcoding of Fresh and Old Museum Specimens. PLoS ONE 6 (7): e21252. https://doi.org/10.1371/journal.pone.0021252
- Smith D, Peay K (2014) Sequence Depth, Not PCR Replication, Improves Ecological Inference from Next Generation DNA Sequencing. PLoS ONE 9 (2): e90234. https://doi.org/10.1371/journal.pone.0090234
- Sunnucks P, Hales DF (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus Sitobion (Hemiptera: Aphididae). Molecular Biology and Evolution 13 (3): 510-524. https://doi.org/10.1093/oxfordjournals.molbev.a025612
- Sweeney B, Battle J, Jackson J, Dapkey T (2011) Can DNA barcodes of stream macroinvertebrates improve descriptions of community structure and water quality? Journal of the North American Benthological Society 30 (1): 195-216. https://doi.org/10.1899/10-016.1
- Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular ecology 21 (8): 2045-50. https://doi.org/10.1111/j.1365-294X.2012.05470.x
- Thomsen PF, Willerslev E (2015) Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. Biological Conservation 183: 4-18. https://doi.org/10.1016/j.biocon.2014.11.019
- Zeale MRK, Butlin RK, Barker GLA, Lees DC, Jones G (2010) Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. Molecular ecology resources 11 (2): 236-44. https://doi.org/10.1111/j.1755-0998.2010.02920.x
- Zhou X, Adamowicz SJ, Jacobus LM, DeWalt RE, Hebert PD (2009) Towards a comprehensive barcode library for arctic life - Ephemeroptera, Plecoptera, and Trichoptera of Churchill, Manitoba, Canada. Frontiers in Zoology 6 (1): 30. https://doi.org/10.1186/1742-9994-6-30
- Zhou X, Jacobus L, DeWalt RE, Adamowicz S, Hebert PN (2010) Ephemeroptera, Plecoptera, and Trichoptera fauna of Churchill (Manitoba, Canada): insights into biodiversity patterns from DNA barcoding. Journal of the North American Benthological Society 29 (3): 814-837. https://doi.org/10.1899/09-121.1
- Zhou X, Frandsen P, Holzenthal R, Beet C, Bennett K, Blahnik R, Bonada N, Cartwright D, Chuluunbat S, Cocks G, Collins G, Waard Jd, Dean J, Flint O, Gonzalez MR, Hausmann A, Hendrich L, Hess M, Hogg I, Kondratieff B, Malicky H, Moriniere J, Morse J, Pauls S, Rinne A, Robinson J, Salokannel J, Shackleton M, Smith B, Stamatakis A,

StClair R, Thomas J, Zamora-Munoz C, Ziesmann T, Kjer K (2016) The Trichoptera barcode initiative: a strategy for generating a species-level tree of life. Barcode of Life Data Systems https://doi.org/10.5883/DS-TBOL

Supplementary materials

Suppl. material 1: Table S1 doi

Authors: Ecaterina Edith Vamos, Vasco Elbrecht, Florian Leese
Data type: Table
Brief description: Primers evaluated in this study
Filename: Tab_S1_meta_primer_v10.xls - Download file (10.00 kb)

Suppl. material 2: Figure S1 doi

Authors: Ecaterina Edith Vamos, Vasco Elbrecht, Florian Leese
Data type: Figure
Brief description: Developed fusion primers for fwh1 and fwh2 on the Illumina high throughput sequencing platform.
Filename: Fig_S1_Fusion_primers.pdf - Download file (24.32 kb)

Suppl. material 3: Figure S2 doi

Authors: Ecaterina Edith Vamos, Vasco Elbrecht, Florian Leese
Data type: Figure
Brief description: Overview of similarity of used inline tags for the fwh1 and fwh2 fusion primers.
Filename: Fig_S2_fwh_combo.pdf - Download file (3.48 MB)

Suppl. material 4: Table S2 doi

Authors: Ecaterina Edith Vamos, Vasco Elbrecht, Florian Leese
Data type: Table
Brief description: Overview of the three Romanian macrozoobenthos sampling sites (Z2, L2, R2).
Filename: Tab_S2.pdf - Download file (269.05 kb)

Suppl. material 5: Figure S3 doi

Authors: Ecaterina Edith Vamos, Vasco Elbrecht, Florian Leese
Data type: Figure
Brief description: Overview of the macroinvertebrates composition of the three sample sites in Romania.
Filename: Fig_S3_sorted.pdf - Download file (628.99 kb)

Suppl. material 6: Table S3 doi

Authors: Ecaterina Edith Vamos, Vasco Elbrecht, Florian Leese

Data type: Table
Brief description: Overview of used tagging combinations for sample multiplexing for sequencing.
Filename: Tab_S3_demultiplexing.xls - Download file (15.00 kb)

Suppl. material 7: Figure S4 doi

Authors: Ecaterina Edith Vamos, Vasco Elbrecht, Florian Leese
Data type: Figure
Brief description: Gradient PCR optimisation for the *fwh* primer sets.
Filename: Fig_S4_gradient.pdf - Download file (193.59 kb)

Suppl. material 8: Script S1 doi

Authors: Ecaterina Edith Vamos, Vasco Elbrecht, Florian Leese
Data type: R Script + faste file (DceM)
Brief description: JAMP metabarcoding pipeline (used R commands) and expected single species mock sample haplotypes (fasta file)
Filename: Scripts S1.zip - Download file (3.31 kb)

Suppl. material 9: Figure S5 doi

Authors: Ecaterina Edith Vamos, Vasco Elbrecht, Florian Leese
Data type: Figure
Brief description: Number of raw sequences obtained for each sample after demultiplexing.
Filename: Fig_S5_demulti.pdf - Download file (652.59 kb)

Suppl. material 10: Table S4 doi

Authors: Ecaterina Edith Vamos, Vasco Elbrecht, Florian Leese
Data type: Table
Brief description: OTU table for the 52 taxa mock samples sequenced with the fwh1 and fwh2 primer set.
Filename: Tab_S4_OTU_table_v2.xls - Download file (434.00 kb)

Suppl. material 11: Figure S6 doi

Authors: Ecaterina Edith Vamos, Vasco Elbrecht, Florian Leese
Data type: Figure
Brief description: Proportion of shared reads between the two replicates for DceM amplified with the fwh1 primer set.
Filename: Fig_S6_shared.pdf - Download file (79.21 kb)

Suppl. material 12: Figure S7 doi

Authors: Ecaterina Edith Vamos, Vasco Elbrecht, Florian Leese
Data type: Figure
Brief description: Sample composition of Romanian macroinvertebrate samples.
Filename: Fig_S7_composition.pdf - Download file (136.06 kb)